

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: TAKAKURA, Hikaru  
MORISHITA, Mio  
YAMAMOTO, Katsuhiko  
MITTA, Masanori  
ASADA, Kiyozo  
TSUNASAWA, Susumu  
KATO, Ikunoshin
- (ii) TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
- (iii) NUMBER OF SEQUENCES: 42
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: Browdy and Neimark  
(B) STREET: 419 Seventh Street N.W., Ste. 300  
(C) CITY: Washington  
(D) STATE: D.C.  
(E) COUNTRY: United States of America  
(F) ZIP: 20004
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER: US 08/894,818  
(B) FILING DATE: 20-MAY-1998  
(C) CLASSIFICATION:
- (vi) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: PCT/JP96/03253  
(B) FILING DATE: 07-NOV-1996
- (vi) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: JP 323285/1995  
(B) FILING DATE: 12-DEC-1995
- (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: Browdy, Roger L.  
(B) REGISTRATION NUMBER: 25,618  
(C) REFERENCE/DOCKET NUMBER: TAKAKURA=1
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (202) 628-5197  
(B) TELEFAX: (202) 737-3528

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 659 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Lys Arg Leu Gly Ala Val Val Leu Ala Leu Val Leu Val Gly  
5 10 15

TECHNOLOGY

Leu	Leu	Ala	Gly	Thr	Ala	Leu	Ala	Ala	Pro	Val	Lys	Pro	Val	Val	20	25	30
Arg	Asn	Asn	Ala	Val	Gln	Gln	Lys	Asn	Tyr	Gly	Leu	Leu	Thr	Pro	35	40	45
Gly	Leu	Phe	Lys	Lys	Val	Gln	Arg	Met	Asn	Trp	Asn	Gln	Glu	Val	50	55	60
Asp	Thr	Val	Ile	Met	Phe	Gly	Ser	Tyr	Gly	Asp	Arg	Asp	Arg	Ala	65	70	75
Val	Lys	Val	Leu	Arg	Leu	Met	Gly	Ala	Gln	Val	Lys	Tyr	Ser	Tyr	80	85	90
Lys	Ile	Ile	Pro	Ala	Val	Ala	Val	Lys	Ile	Lys	Ala	Arg	Asp	Leu	95	100	105
Leu	Leu	Ile	Ala	Gly	Met	Ile	Asp	Thr	Gly	Tyr	Phe	Gly	Asn	Thr	110	115	120
Arg	Val	Ser	Gly	Ile	Lys	Phe	Ile	Gln	Glu	Asp	Tyr	Lys	Val	Gln	125	130	135
Val	Asp	Asp	Ala	Thr	Ser	Val	Ser	Gln	Ile	Gly	Ala	Asp	Thr	Val	140	145	150
Trp	Asn	Ser	Leu	Gly	Tyr	Asp	Gly	Ser	Gly	Val	Val	Val	Ala	Ile	155	160	165
Val	Asp	Thr	Gly	Ile	Asp	Ala	Asn	His	Pro	Asp	Leu	Lys	Gly	Lys	170	175	180
Val	Ile	Gly	Trp	Tyr	Asp	Ala	Val	Asn	Gly	Arg	Ser	Thr	Pro	Tyr	185	190	195
Asp	Asp	Gln	Gly	His	Gly	Thr	His	Val	Ala	Gly	Ile	Val	Ala	Gly	200	205	210
Thr	Gly	Ser	Val	Asn	Ser	Gln	Tyr	Ile	Gly	Val	Ala	Pro	Gly	Ala	215	220	225
Lys	Leu	Val	Gly	Val	Lys	Val	Leu	Gly	Ala	Asp	Gly	Ser	Gly	Ser	230	235	240
Val	Ser	Thr	Ile	Ile	Ala	Gly	Val	Asp	Trp	Val	Val	Gln	Asn	Lys	245	250	255
Asp	Lys	Tyr	Gly	Ile	Arg	Val	Ile	Asn	Leu	Ser	Leu	Gly	Ser	Ser	260	265	270
Gln	Ser	Ser	Asp	Gly	Thr	Asp	Ser	Leu	Ser	Gln	Ala	Val	Asn	Asn	275	280	285
Ala	Trp	Asp	Ala	Gly	Ile	Val	Val	Cys	Val	Ala	Ala	Gly	Asn	Ser	290	295	300
Gly	Pro	Asn	Thr	Tyr	Thr	Val	Gly	Ser	Pro	Ala	Ala	Ala	Ser	Lys	305	310	315
Val	Ile	Thr	Val	Gly	Ala	Val	Asp	Ser	Asn	Asp	Asn	Ile	Ala	Ser	320	325	330
Phe	Ser	Ser	Arg	Gly	Pro	Thr	Ala	Asp	Gly	Arg	Leu	Lys	Pro	Glu	335	340	345

034530

Val	Val	Ala	Pro	Gly	Val	Asp	Ile	Ile	Ala	Pro	Arg	Ala	Ser	Gly	350	355	360
Thr	Ser	Met	Gly	Thr	Pro	Ile	Asn	Asp	Tyr	Tyr	Thr	Lys	Ala	Ser	365	370	375
Gly	Thr	Ser	Met	Ala	Thr	Pro	His	Val	Ser	Gly	Val	Gly	Ala	Leu	380	385	390
Ile	Leu	Gln	Ala	His	Pro	Ser	Trp	Thr	Pro	Asp	Lys	Val	Lys	Thr	395	400	405
Ala	Leu	Ile	Glu	Thr	Ala	Asp	Ile	Val	Ala	Pro	Lys	Glu	Ile	Ala	410	415	420
Asp	Ile	Ala	Tyr	Gly	Ala	Gly	Arg	Val	Asn	Val	Tyr	Lys	Ala	Ile	425	430	435
Lys	Tyr	Asp	Asp	Tyr	Ala	Lys	Leu	Thr	Phe	Thr	Gly	Ser	Val	Ala	440	445	450
Asp	Lys	Gly	Ser	Ala	Thr	His	Thr	Phe	Asp	Val	Ser	Gly	Ala	Thr	455	460	465
Phe	Val	Thr	Ala	Thr	Leu	Tyr	Trp	Asp	Thr	Gly	Ser	Ser	Asp	Ile	470	475	480
Asp	Leu	Tyr	Leu	Tyr	Asp	Pro	Asn	Gly	Asn	Glu	Val	Asp	Tyr	Ser	485	490	495
Tyr	Thr	Ala	Tyr	Tyr	Gly	Phe	Glu	Lys	Val	Gly	Tyr	Tyr	Asn	Pro	500	505	510
Thr	Ala	Gly	Thr	Trp	Thr	Val	Lys	Val	Val	Ser	Tyr	Lys	Gly	Ala	515	520	525
Ala	Asn	Tyr	Gln	Val	Asp	Val	Val	Ser	Asp	Gly	Ser	Leu	Ser	Gln	530	535	540
Ser	Gly	Gly	Gly	Asn	Pro	Asn	Pro	Asn	Pro	Asn	Pro	Asn	Pro	Thr	545	550	555
Pro	Thr	Thr	Asp	Thr	Gln	Thr	Phe	Thr	Gly	Ser	Val	Asn	Asp	Tyr	560	565	570
Trp	Asp	Thr	Ser	Asp	Thr	Phe	Thr	Met	Asn	Val	Asn	Ser	Gly	Ala	575	580	585
Thr	Lys	Ile	Thr	Gly	Asp	Leu	Thr	Phe	Asp	Thr	Ser	Tyr	Asn	Asp	590	595	600
Leu	Asp	Leu	Tyr	Leu	Tyr	Asp	Pro	Asn	Gly	Asn	Leu	Val	Asp	Arg	605	610	615
Ser	Thr	Ser	Ser	Asn	Ser	Tyr	Glu	His	Val	Glu	Tyr	Ala	Asn	Pro	620	625	630
Ala	Pro	Gly	Thr	Trp	Thr	Phe	Leu	Val	Tyr	Ala	Tyr	Ser	Thr	Tyr	635	640	645
Gly	Trp	Ala	Asp	Tyr	Gln	Leu	Lys	Ala	Val	Val	Tyr	Tyr	Gly		650	655	

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

034155-042401

(A) LENGTH: 1977 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGAAGAGGT	TAGGTGCTGT	GGTGCTGGCA	CTGGTGCTCG	TGGGTCTTCT	GGCCGGAACG	60
GCCCTTGCGG	CACCCGTAAA	ACCGGTTGTC	AGGAACAACG	CGGTTCAGCA	GAAGAACTAC	120
GGACTGCTGA	CCCCGGGACT	GTTCAAGAAA	GTCCAGAGGA	TGAACTGGAA	CCAGGAAGTG	180
GACACCGTCA	TAATGTTCTG	GAGCTACGGA	GACAGGGACA	GGGCGGTTAA	GGTACTGAGG	240
CTCATGGGCG	CCCAGGTCAA	GTACTCCTAC	AAGATAATCC	CTGCTGTCGC	GGTTAAAATA	300
AAGGCCAGGG	ACCTTCTGCT	GATCGCGGGC	ATGATAGACA	CGGGTTACTT	CGGTAACACA	360
AGGGTCTCGG	GCATAAAGTT	CATACAGGAG	GATTACAAGG	TTCAGGTTGA	CGACGCCACT	420
TCCGTCTCCC	AGATAGGGGC	CGATACCGTC	TGGAAC TCCC	TCGGCTACGA	CGGAAGCGGT	480
GTGGTG GTTG	CCATCGTCGA	TACGGGTATA	GACGCGAACC	ACCCCGATCT	GAAGGGCAAG	540
GTCATAGGCT	GGTACGACGC	CGTCAACGGC	AGGTCGACCC	CCTACGATGA	CCAGGGACAC	600
GGAACCCACG	TTGCGGGTAT	CGTTGCCGGA	ACCGGCAGCG	TTAACTCCCA	GTACATAGGC	660
GTCGCCCCCG	GCGCGAAGCT	CGTCGGCGTC	AAGGTTCTCG	GTGCCGACGG	TTCGGGAAGC	720
GTCTCCACCA	TCATCGCGGG	TGTTGACTGG	GTCTGCCAGA	ACAAGGACAA	GTACGGGATA	780
AGGGTCATCA	ACCTCTCCCT	CGGCTCCTCC	CAGAGCTCCG	ACGGAACCGA	CTCCCTCAGT	840
CAGGCCGTCA	ACAACGCCTG	GGACGCCGGT	ATAGTAGTCT	GCGTCGCCGC	CGGCAACAGC	900
GGGCCGAACA	CCTACACCGT	CGGCTCACCC	GCCGCCGCGA	GCAAGGTCAT	AACCGTCGGT	960
GCAGTTGACA	GCAACGACAA	CATCGCCAGC	TTCTCCAGCA	GGGGACCGAC	CGCGGACGGA	1020
AGGCTCAAGC	CGGAAGTCGT	CGCCCCCGGC	GTTGACATCA	TAGCCCCGCG	CGCCAGCGGA	1080
ACCAGCATGG	GCACCCCGAT	AAACGACTAC	TACACCAAGG	CCTCTGGAAC	CAGCATGGCC	1140
ACCCCGCACG	TTTCGGGCGT	TGGCGCGCTC	ATCCTCCAGG	CCCACCCGAG	CTGGACCCCG	1200
GACAAGGTGA	AGACCGCCCT	CATCGAGACC	GCCGACATAG	TCGCCCCCAA	GGAGATAGCG	1260
GACATCGCCT	ACGGTGCGGG	TAGGGTGAAC	GTCTACAAGG	CCATCAAGTA	CGACGACTAC	1320
GCCAAGCTCA	CCTTCACCGG	CTCCGTCGCC	GACAAGGGAA	GCGCCACCCA	CACCTTCGAC	1380
GTCAGCGGCG	CCACCTTCGT	GACCGCCACC	CTCTACTGGG	ACACGGGCTC	GAGCGACATC	1440
GACCTCTACC	TCTACGACCC	CAACGGGAAC	GAGGTTGACT	ACTCCTACAC	CGCCTACTAC	1500
GGCTTCGAGA	AGGTCGGCTA	CTACAACCCG	ACCGCCGGAA	CCTGGACGGT	CAAGGTCGTC	1560
AGCTACAAGG	GCGCGGCGAA	CTACCAGGTC	GACGTCGTCA	GCGACGGGAG	CCTCAGCCAG	1620
TCCGGCGGCG	GCAACCCGAA	TCCAAACCCC	AACCCGAACC	CAACCCCGAC	CACCGACACC	1680
CAGACCTTCA	CCGTTTCCGT	TAACGACTAC	TGGGACACCA	GCGACACCTT	CACCATGAAC	1740

GTCAACAGCG GTGCCACCAA GATAACCGGT GACCTGACCT TCGATACTTC CTACAACGAC 1800  
CTCGACCTCT ACCTCTACGA CCCCACGGC AACCTCGTTG ACAGGTCCAC GTCGAGCAAC 1860  
AGCTACGAGC ACGTCGAGTA CGCCAACCCC GCCCCGGGAA CCTGGACGTT CCTCGTCTAC 1920  
GCCTACAGCA CCTACGGCTG GGCGGACTAC CAGCTCAAGG CCGTCGTCTA CTACGGG 1977

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 522 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (D) OTHER INFORMATION:/note= Xaa at position 428 is Gly or Val.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ala	Glu	Leu	Glu	Gly	Leu	Asp	Glu	Ser	Ala	Ala	Gln	Val	Met	Ala	
				5					10					15	
Thr	Tyr	Val	Trp	Asn	Leu	Gly	Tyr	Asp	Gly	Ser	Gly	Ile	Thr	Ile	
				20					25					30	
Gly	Ile	Ile	Asp	Thr	Gly	Ile	Asp	Ala	Ser	His	Pro	Asp	Leu	Gln	
				35					40					45	
Gly	Lys	Val	Ile	Gly	Trp	Val	Asp	Phe	Val	Asn	Gly	Arg	Ser	Tyr	
				50					55					60	
Pro	Tyr	Asp	Asp	His	Gly	His	Gly	Thr	His	Val	Ala	Ser	Ile	Ala	
				65					70					75	
Ala	Gly	Thr	Gly	Ala	Ala	Ser	Asn	Gly	Lys	Tyr	Lys	Gly	Met	Ala	
				80					85					90	
Pro	Gly	Ala	Lys	Leu	Ala	Gly	Ile	Lys	Val	Leu	Gly	Ala	Asp	Gly	
				95					100					105	
Ser	Gly	Ser	Ile	Ser	Thr	Ile	Ile	Lys	Gly	Val	Glu	Trp	Ala	Val	
				110					115					120	
Asp	Asn	Lys	Asp	Lys	Tyr	Gly	Ile	Lys	Val	Ile	Asn	Leu	Ser	Leu	
				125					130					135	
Gly	Ser	Ser	Gln	Ser	Ser	Asp	Gly	Thr	Asp	Ala	Leu	Ser	Gln	Ala	
				140					145					150	
Val	Asn	Ala	Ala	Trp	Asp	Ala	Gly	Leu	Val	Val	Val	Val	Ala	Ala	
				155					160					165	
Gly	Asn	Ser	Gly	Pro	Asn	Lys	Tyr	Thr	Ile	Gly	Ser	Pro	Ala	Ala	
				170					175					180	
Ala	Ser	Lys	Val	Ile	Thr	Val	Gly	Ala	Val	Asp	Lys	Tyr	Asp	Val	
				185					190					195	
Ile	Thr	Ser	Phe	Ser	Ser	Arg	Gly	Pro	Thr	Ala	Asp	Gly	Arg	Leu	
				200					205					210	

004240" E55F4360

Lys	Pro	Glu	Val	Val	Ala	Pro	Gly	Asn	Trp	Ile	Ile	Ala	Ala	Arg
				215					220					225
Ala	Ser	Gly	Thr	Ser	Met	Gly	Gln	Pro	Ile	Asn	Asp	Tyr	Tyr	Thr
				230					235					240
Ala	Ala	Pro	Gly	Thr	Ser	Met	Ala	Thr	Pro	His	Val	Ala	Gly	Ile
				245					250					255
Ala	Ala	Leu	Leu	Leu	Gln	Ala	His	Pro	Ser	Trp	Thr	Pro	Asp	Lys
				260					265					270
Val	Lys	Thr	Ala	Leu	Ile	Glu	Thr	Ala	Asp	Ile	Val	Lys	Pro	Asp
				275					280					285
Glu	Ile	Ala	Asp	Ile	Ala	Tyr	Gly	Ala	Gly	Arg	Val	Asn	Ala	Tyr
				290					295					300
Lys	Ala	Ile	Asn	Tyr	Asp	Asn	Tyr	Ala	Lys	Leu	Val	Phe	Thr	Gly
				305					310					315
Tyr	Val	Ala	Asn	Lys	Gly	Ser	Gln	Thr	His	Gln	Phe	Val	Ile	Ser
				320					325					330
Gly	Ala	Ser	Phe	Val	Thr	Ala	Thr	Leu	Tyr	Trp	Asp	Asn	Ala	Asn
				335					340					345
Ser	Asp	Leu	Asp	Leu	Tyr	Leu	Tyr	Asp	Pro	Asn	Gly	Asn	Gln	Val
				350					355					360
Asp	Tyr	Ser	Tyr	Thr	Ala	Tyr	Tyr	Gly	Phe	Glu	Lys	Val	Gly	Tyr
				365					370					375
Tyr	Asn	Pro	Thr	Asp	Gly	Thr	Trp	Thr	Ile	Lys	Val	Val	Ser	Tyr
				380					385					390
Ser	Gly	Ser	Ala	Asn	Tyr	Gln	Val	Asp	Val	Val	Ser	Asp	Gly	Ser
				395					400					405
Leu	Ser	Gln	Pro	Gly	Ser	Ser	Pro	Ser	Pro	Gln	Pro	Glu	Pro	Thr
				410					415					420
Val	Asp	Ala	Lys	Thr	Phe	Gln	Xaa	Ser	Asp	His	Tyr	Tyr	Tyr	Asp
				425					430					435
Arg	Ser	Asp	Thr	Phe	Thr	Met	Thr	Val	Asn	Ser	Gly	Ala	Thr	Lys
				440					445					450
Ile	Thr	Gly	Asp	Leu	Val	Phe	Asp	Thr	Ser	Tyr	His	Asp	Leu	Asp
				455					460					465
Leu	Tyr	Leu	Tyr	Asp	Pro	Asn	Gln	Lys	Leu	Val	Asp	Arg	Ser	Glu
				470					475					480
Ser	Pro	Asn	Ser	Tyr	Glu	His	Val	Glu	Tyr	Leu	Thr	Pro	Ala	Pro
				485					490					495
Gly	Thr	Trp	Tyr	Phe	Leu	Val	Tyr	Ala	Tyr	Tyr	Thr	Tyr	Gly	Trp
				500					505					510
Ala	Tyr	Tyr	Glu	Leu	Thr	Ala	Lys	Val	Tyr	Tyr	Gly			
				515					520					

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

00440" E55F4360

(A) LENGTH: 1566 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(ix) FEATURE:

(D) OTHER INFORMATION: /note= N at position 1283 is G or T.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GCAGAATTAG AAGGACTGGA TGAGTCTGCA GCTCAAGTTA TGGCAACTTA CGTTTGGAAAC	60
TTGGGATATG ATGGTTCCTGG AATCACAATA GGAATAATTG AACTTGGAAT TGACGCTTCT	120
CATCCAGATC TCCAAGGAAA AGTAATTGGG TGGGTAGATT TTGTCAATGG TAGGAGTTAT	180
CCATACGATG ACCATGGACA TGGAACTCAT GTAGCTTCAA TAGCAGCTGG TACTGGAGCA	240
GCAAGTAATG GCAAGTACAA GGAATGGCT CCAGGAGCTA AGCTGGCGGG AATTAAGGTT	300
CTAGGTGCCG ATGGTTCCTGG AAGCATATCT ACTATAATTA AGGGAGTTGA GTGGGCCGTT	360
GATAACAAAG ATAAGTACGG AATTAAGGTC ATTAATCTTT CTCTTGTTTC AAGCCAGAGC	420
TCAGATGGTA CTGACGCTCT AAGTCAGGCT GTTAATGCAG CGTGGGATGC TGGATTAGTT	480
GTTGTGGTTG CCGCTGGAAA CAGTGGACCT AACAAGTATA CAATCGGTTT TCCAGCAGCT	540
GCAAGCAAAG TTATTACAGT TGGAGCCGTT GACAAGTATG ATGTTATAAC AAGCTTCTCA	600
AGCAGAGGGC CAACTGCAGA CGGCAGGCTT AAGCCTGAGG TTGTTGCTCC AGGAAACTGG	660
ATAATTGCTG CCAGAGCAAG TGGAACTAGC ATGGGTCAAC CAATTAATGA CTATTACACA	720
GCAGCTCTTG GGACATCAAT GGCAACTCCT CACGTAGCTG GTATTGCAGC CCTCTTGCTC	780
CAAGCACACC CGAGCTGGAC TCCAGACAAA GTAAAAACAG CCCTCATAGA AACTGCTGAT	840
ATCGTAAAGC CAGATGAAAT AGCCGATATA GCCTACGGTG CAGGTAGGGT TAATGCATAC	900
AAGGCTATAA ACTACGATAA CTATGCAAAG CTAGTGTTCA CTGGATATGT TGCCAACAAA	960
GGCAGCCAAA CTCACCAGTT CGTTATTAGC GGAGCTTCGT TCGTAACTGC CACATTATAC	1020
TGGGACAATG CCAATAGCGA CCTTGATCTT TACCTCTACG ATCCCAATGG AAACCAGGTT	1080
GACTACTCTT ACACCGCCTA CTATGGATTC GAAAAGGTTG GTTATTACAA CCCAACTGAT	1140
GGAACATGGA CAATTAAGGT TGTAAGCTAC AGCGGAAGTG CAAACTATCA AGTAGATGTG	1200
GTAAGTGATG GTTCCCTTTC ACAGCCTGGA AGTTCACCAT CTCCACAACC AGAACCAACA	1260
GTAGACGCAA AGACGTTCCA AGNATCCGAT CACTACTACT ATGACAGGAG CGACACCTTT	1320
ACAATGACCG TTAAGTCTGG GGCTACAAAG ATTACTGGAG ACCTAGTGTT TGACACAAGC	1380
TACCATGATC TTGACCTTTA CCTCTACGAT CCTAACCAGA AGCTTGTTAG TAGATCGGAG	1440
AGTCCCAACA GCTACGAACA CGTAGAATAC TTAACCCCG CCCCAGGAAC CTGGTACTTC	1500
CTAGTATATG CCTACTACAC TTACGGTTGG GCTTACTACG AGCTGACGGC TAAAGTTTAT	1560
TATGGC	1566

03041533 042401

(2) INFORMATION FOR SEQ ID NO:5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 659 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Lys	Gly	Leu	Lys	Ala	Leu	Ile	Leu	Val	Ile	Leu	Val	Leu	Gly	5	10	15
Leu	Val	Val	Gly	Ser	Val	Ala	Ala	Ala	Pro	Glu	Lys	Lys	Val	Glu	20	25	30
Gln	Val	Arg	Asn	Val	Glu	Lys	Asn	Tyr	Gly	Leu	Leu	Thr	Pro	Gly	35	40	45
Leu	Phe	Arg	Lys	Ile	Gln	Lys	Leu	Asn	Pro	Asn	Glu	Glu	Ile	Ser	50	55	60
Thr	Val	Ile	Val	Phe	Glu	Asn	His	Arg	Glu	Lys	Glu	Ile	Ala	Val	65	70	75
Arg	Val	Leu	Glu	Leu	Met	Gly	Ala	Lys	Val	Arg	Tyr	Val	Tyr	His	80	85	90
Ile	Ile	Pro	Ala	Ile	Ala	Ala	Asp	Leu	Lys	Val	Arg	Asp	Leu	Leu	95	100	105
Val	Ile	Ser	Gly	Leu	Thr	Gly	Gly	Lys	Ala	Lys	Leu	Ser	Gly	Val	110	115	120
Arg	Phe	Ile	Gln	Glu	Asp	Tyr	Lys	Val	Thr	Val	Ser	Ala	Glu	Leu	125	130	135
Glu	Gly	Leu	Asp	Glu	Ser	Ala	Ala	Gln	Val	Met	Ala	Thr	Tyr	Val	140	145	150
Trp	Asn	Leu	Gly	Tyr	Asp	Gly	Ser	Gly	Ile	Thr	Ile	Gly	Ile	Ile	155	160	165
Asp	Thr	Gly	Ile	Asp	Ala	Ser	His	Pro	Asp	Leu	Gln	Gly	Lys	Val	170	175	180
Ile	Gly	Trp	Val	Asp	Phe	Val	Asn	Gly	Arg	Ser	Tyr	Pro	Tyr	Asp	185	190	195
Asp	His	Gly	His	Gly	Thr	His	Val	Ala	Ser	Ile	Ala	Ala	Gly	Thr	200	205	210
Gly	Ala	Ala	Ser	Asn	Gly	Lys	Tyr	Lys	Gly	Met	Ala	Pro	Gly	Ala	215	220	225
Lys	Leu	Ala	Gly	Ile	Lys	Val	Leu	Gly	Ala	Asp	Gly	Ser	Gly	Ser	230	235	240
Ile	Ser	Thr	Ile	Ile	Lys	Gly	Val	Glu	Trp	Ala	Val	Asp	Asn	Lys	245	250	255
Asp	Lys	Tyr	Gly	Ile	Lys	Val	Ile	Asn	Leu	Ser	Leu	Gly	Ser	Ser	260	265	270

093455304404



Gln	Ser	Ser	Asp	Gly	Thr	Asp	Ser	Leu	Ser	Gln	Ala	Val	Asn	Asn	
				275					280					285	
Ala	Trp	Asp	Ala	Gly	Ile	Val	Val	Cys	Val	Ala	Ala	Gly	Asn	Ser	
				290					295					300	
Gly	Pro	Asn	Thr	Tyr	Thr	Val	Gly	Ser	Pro	Ala	Ala	Ala	Ser	Lys	
				305					310					315	
Val	Ile	Thr	Val	Gly	Ala	Val	Asp	Ser	Asn	Asp	Asn	Ile	Ala	Ser	
				320					325					330	
Phe	Ser	Ser	Arg	Gly	Pro	Thr	Ala	Asp	Gly	Arg	Leu	Lys	Pro	Glu	
				335					340					345	
Val	Val	Ala	Pro	Gly	Val	Asp	Ile	Ile	Ala	Pro	Arg	Ala	Ser	Gly	
				350					355					360	
Thr	Ser	Met	Gly	Thr	Pro	Ile	Asn	Asp	Tyr	Tyr	Thr	Lys	Ala	Ser	
				365					370					375	
Gly	Thr	Ser	Met	Ala	Thr	Pro	His	Val	Ser	Gly	Val	Gly	Ala	Leu	
				380					385					390	
Ile	Leu	Gln	Ala	His	Pro	Ser	Trp	Thr	Pro	Asp	Lys	Val	Lys	Thr	
				395					400					405	
Ala	Leu	Ile	Glu	Thr	Ala	Asp	Ile	Val	Ala	Pro	Lys	Glu	Ile	Ala	
				410					415					420	
Asp	Ile	Ala	Tyr	Gly	Ala	Gly	Arg	Val	Asn	Val	Tyr	Lys	Ala	Ile	
				425					430					435	
Lys	Tyr	Asp	Asp	Tyr	Ala	Lys	Leu	Thr	Phe	Thr	Gly	Ser	Val	Ala	
				440					445					450	
Asp	Lys	Gly	Ser	Ala	Thr	His	Thr	Phe	Asp	Val	Ser	Gly	Ala	Thr	
				455					460					465	
Phe	Val	Thr	Ala	Thr	Leu	Tyr	Trp	Asp	Thr	Gly	Ser	Ser	Asp	Ile	
				470					475					480	
Asp	Leu	Tyr	Leu	Tyr	Asp	Pro	Asn	Gly	Asn	Glu	Val	Asp	Tyr	Ser	
				485					490					495	
Tyr	Thr	Ala	Tyr	Tyr	Gly	Phe	Glu	Lys	Val	Gly	Tyr	Tyr	Asn	Pro	
				500					505					510	
Thr	Ala	Gly	Thr	Trp	Thr	Val	Lys	Val	Val	Ser	Tyr	Lys	Gly	Ala	
				515					520					525	
Ala	Asn	Tyr	Gln	Val	Asp	Val	Val	Ser	Asp	Gly	Ser	Leu	Ser	Gln	
				530					535					540	
Ser	Gly	Gly	Gly	Asn	Pro	Asn	Pro	Asn	Pro	Asn	Pro	Asn	Pro	Thr	
				545					550					555	
Pro	Thr	Thr	Asp	Thr	Gln	Thr	Phe	Thr	Gly	Ser	Val	Asn	Asp	Tyr	
				560					565					570	
Trp	Asp	Thr	Ser	Asp	Thr	Phe	Thr	Met	Asn	Val	Asn	Ser	Gly	Ala	
				575					580					585	
Thr	Lys	Ile	Thr	Gly	Asp	Leu	Thr	Phe	Asp	Thr	Ser	Tyr	Asn	Asp	
				590					595					600	

TC4240" ESSF450

Leu Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Leu Val Asp Arg  
605 610 615  
Ser Thr Ser Ser Asn Ser Tyr Glu His Val Glu Tyr Ala Asn Pro  
620 625 630  
Ala Pro Gly Thr Trp Thr Phe Leu Val Tyr Ala Tyr Ser Thr Tyr  
635 640 645  
Gly Trp Ala Asp Tyr Gln Leu Lys Ala Val Val Tyr Tyr Gly  
650 655

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1977 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGAAGGGGC	TGAAAGCTCT	CATATTAGTG	ATTTTAGTTC	TAGGTTTGGT	AGTAGGGAGC	60
GTAGCGGCAG	CTCCAGAGAA	GAAAGTTGAA	CAAGTAAGAA	ATGTTGAGAA	GAACATGCTT	120
CTGCTAACGC	CAGGACTGTT	CAGAAAAATT	CAAAAATTGA	ATCCTAACGA	GGAAATCAGC	180
ACAGTAATTG	TATTTGAAAA	CCATAGGGAA	AAAGAAATTG	CAGTAAGAGT	TCTTGAGTTA	240
ATGGGTGCAA	AAGTTAGGTA	TGTGTACCAT	ATTATACCCG	CAATAGCTGC	CGATCTTAAG	300
GTTAGAGACT	TACTAGTCAT	CTCAGGTTTA	ACAGGGGGTA	AAGCTAAGCT	TTCAGGTGTT	360
AGGTTTATCC	AGGAAGACTA	CAAAGTTACA	GTTTCAGCAG	AATTAGAAGG	ACTGGATGAG	420
TCTGCAGCTC	AAGTTATGGC	AACTTACGTT	TGGAACCTTG	GATATGATGG	TTCTGGAATC	480
ACAATAGGAA	TAATTGACAC	TGGAATTGAC	GCTTCTCATC	CAGATCTCCA	AGGAAAAGTA	540
ATTGGGTGGG	TAGATTTTGT	CAATGGTAGG	AGTTATCCAT	ACGATGACCA	TGGACATGGA	600
ACTCATGTAG	CTTCAATAGC	AGCTGGTACT	GGAGCAGCAA	GTAATGGCAA	GTACAAGGGA	660
ATGGCTCCAG	GAGCTAAGCT	GGCGGGAATT	AAGGTCTAG	GTGCCGATGG	TTCTGGAAGC	720
ATATCTACTA	TAATTAAGGG	AGTTGAGTGG	GCCGTTGATA	ACAAAGATAA	GTACGGAATT	780
AAGGTCATTA	ATCTTTCTCT	TGGTTCAAGC	CAGAGCTCCG	ACGGAACCGA	CTCCCTCAGT	840
CAGGCCGTCA	ACAACGCCTG	GGACGCCGGT	ATAGTAGTCT	GCGTCGCCGC	CGGCAACAGC	900
GGGCCGAACA	CCTACACCGT	CGGCTCACCC	GCCGCCGCGA	GCAAGGTCAT	AACCGTCGGT	960
GCAGTTGACA	GCAACGACAA	CATCGCCAGC	TTCTCCAGCA	GGGACCGAC	CGCGGACGGA	1020
AGGCTCAAGC	CGGAAGTCGT	CGCCCCCGGC	GTTGACATCA	TAGCCCCGCG	CGCCAGCGGA	1080
ACCAGCATGG	GCACCCCGAT	AAACGACTAC	TACACCAAGG	CCTCTGGAAC	CAGCATGGCC	1140
ACCCCGCACG	TTTCGGGCGT	TGGCGCGCTC	ATCCTCCAGG	CCCACCCGAG	CTGGACCCCG	1200
GACAAGGTGA	AGACCGCCCT	CATCGAGACC	GCCGACATAG	TCGCCCCCAA	GGAGATAGCG	1260
GACATCGCCT	ACGGTGCGGG	TAGGGTGAAC	GTCTACAAGG	CCATCAAGTA	CGACGACTAC	1320

GCCAAGCTCA	CCTTCACCGG	CTCCGTCGCC	GACAAGGGAA	GCGCCACCCA	CACCTTCGAC	1380
GTCAGCGGCG	CCACCTTCGT	GACCGCCACC	CTCTACTGGG	ACACGGGCTC	GAGCGACATC	1440
GACCTCTACC	TCTACGACCC	CAACGGGAAC	GAGGTTGACT	ACTCCTACAC	CGCCTACTAC	1500
GGCTTCGAGA	AGGTTCGGCTA	CTACAACCCG	ACCGCCGGAA	CCTGGACGGT	CAAGGTCGTC	1560
AGCTACAAGG	GCGCGGCGAA	CTACCAGGTC	GACGTCGTCA	GCGACGGGAG	CCTCAGCCAG	1620
TCCGGCGGCG	GCAACCCGAA	TCCAAACCCC	AACCCGAACC	CAACCCCGAC	CACCGACACC	1680
CAGACCTTCA	CCGTTTCCGT	TAACGACTAC	TGGGACACCA	GCGACACCTT	CACCATGAAC	1740
GTCAACAGCG	GTGCCACCAA	GATAACCGGT	GACCTGACCT	TCGATACTTC	CTACAACGAC	1800
CTCGACCTCT	ACCTCTACGA	CCCCAACGGC	AACCTCGTTG	ACAGGTCCAC	GTGAGCAAC	1860
AGCTACGAGC	ACGTCGAGTA	CGCCAACCCC	GCCCCGGGAA	CCTGGACGTT	CCTCGTCTAC	1920
GCCTACAGCA	CCTACGGCTG	GGCGGACTAC	CAGCTCAAGG	CCGTCGTCTA	CTACGGG	1977

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4765 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTTAAATTAT	AAGATATAAT	CACTCCGAGT	GATGAGTAAG	ATACATCATT	ACAGTCCCAA	60
AATGTTTATA	ATTGGAACGC	AGTGAATATA	CAAATGAAT	ATAACCTCGG	AGGTGACTGT	120
AGAATGAATA	AGAAGGGACT	TACTGTGCTA	TTTATAGCGA	TAATGCTCCT	TTCAGTAGTT	180
CCAGTGCACT	TTGTGTCCGC	AGAAACACCA	CCGGTTAGTT	CAGAAAATTC	AACAACCTTCT	240
ATACTCCCTA	ACCAACAAGT	TGTGACAAAA	GAAGTTTCAC	AAGCGGCGCT	TAATGCTATA	300
ATGAAAGGAC	AACCCAACAT	GGTTCTTATA	ATCAAGACTA	AGGAAGGCAA	ACTTGAAGAG	360
GCAAAAACCG	AGCTTGAAAA	GCTAGGTGCA	GAGATTCTTG	ACGAAAATAG	AGTTCTTAAC	420
ATGTTGCTAG	TTAAGATTAA	GCCTGAGAAA	GTAAAGAGC	TCAACTATAT	CTCATCTCTT	480
GAAAAAGCCT	GGCTTAACAG	AGAAGTTAAG	CTTTCCCCTC	CAATTGTCGA	AAAGGACGTC	540
AAGACTAAGG	AGCCCTCCCT	AGAACCAAAA	ATGTATAACA	GCACCTGGGT	AATTAATGCT	600
CTCCAGTTCA	TCCAGGAATT	TGGATATGAT	GGTAGTGGTG	TTGTTGTTGC	AGTACTTGAC	660
ACGGGAGTTG	ATCCGAACCA	TCCTTTCTTG	AGCATAACTC	CAGATGGACG	CAGGAAAATT	720
ATAGAATGGA	AGGATTTTAC	AGACGAGGGA	TTCGTGGATA	CATCATTCAG	CTTTAGCAAG	780
GTTGTAAATG	GGACTCTTAT	AATTAACACA	ACATTCCAAG	TGGCCTCAGG	TCTCACGCTG	840
AATGAATCGA	CAGGACTTAT	GGAATACGTT	GTAAAGACTG	TTTACGTGAG	CAATGTGACC	900
ATTGGAAATA	TCACTTCTGC	TAATGGCATC	TATCACTTCG	GCCTGCTCCC	AGAAAGATAC	960

TTCGACTTAA ACTTCGATGG TGATCAAGAG GACTTCTATC CTGTCTTATT AGTTAACTCC 1020  
ACTGGCAATG GTTATGACAT TGCATATGTG GATACTGACC TTGACTACGA CTTACCCGAC 1080  
GAAGTTCCAC TTGGCCAGTA CAACGTTACT TATGATGTTG CTGTTTTTAG CTACTACTAC 1140  
GGTCCTCTCA ACTACGTGCT TGCAGAAATA GATCCTAACG GAGAATATGC AGTATTTGGG 1200  
TGGGATGGTC ACGGTCACGG AACTCACGTA GCTGGAAGTGTG TTGCTGGTTA CGACAGCAAC 1260  
AATGATGCTT GGGATTGGCT CAGTATGTAC TCTGGTGAAT GGAAGTGTG CTCAAGACTC 1320  
TATGGTTGGG ATTATACGAA CGTTACCACA GACACCGTGC AGGGTGTGTC TCCAGGTGCC 1380  
CAAATAATGG CAATAAGAGT TCTTAGGAGT GATGGACGGG GTAGCATGTG GGATATTATA 1440  
GAAGGTATGA CATAACGAGC AACCCATGGT GCAGACGTTA TAAGCATGAG TCTCGGTGGA 1500  
AATGCTCCAT ACTTAGATGG TACTGATCCA GAAAGCGTTG CTGTGGATGA GCTTACCGAA 1560  
AAGTACGGTG TTGTATTCGT AATAGCTGCA GGAAATGAAG GTCCTGGCAT TAACATCGTT 1620  
GGAAGTCCTG GTGTTGCAAC AAAGGCAATA ACTGTTGGAG CTGCTGCAGT GCCCATTAAC 1680  
GTTGGAGTTT ATGTTTCCCA AGCACTTGGA TATCCTGATT ACTATGGATT CTATTACTTC 1740  
CCCGCTACA CAAACGTTAG AATAGCATTC TTCTCAAGCA GAGGGCCGAG AATAGATGGT 1800  
GAAATAAAAC CCAATGTAGT GGCTCCAGGT TACGGAATTT ACTCATCCCT GCCGATGTGG 1860  
ATTGGCGGAG CTGACTTCAT GTCTGGAAGT TCGATGGCTA CTCCACATGT CAGCGGTGTC 1920  
GTTGCACTCC TCATAAGCGG GGCAAAGGCC GAGGGAATAT ACTACAATCC AGATATAATT 1980  
AAGAAGGTTT TTGAGAGCGG TGCAACCTGG CTTGAGGGAG ATCCATATAC TGGGCAGAAG 2040  
TACACTGAGC TTGACCAAGG TCATGGTCTT GTTAACGTTA CCAAGTCCTG GGAAATCCTT 2100  
AAGGCTATAA ACGGCACCAC TCTCCCAATT GTTGATCACT GGCAGACAA GTCCTACAGC 2160  
GACTTTGCGG AGTACTTGGG TGTGGACGTT ATAAGAGGTC TCTACGCAAG GAACTCTATA 2220  
CCTGACATTG TCGAGTGGCA CATTAAGTAC GTAGGGGACA CGGAGTACAG AACTTTTGAG 2280  
ATCTATGCAA CTGAGCCATG GATTAAGCCT TTTGTCAGTG GAAGTGTAAT TCTAGAGAAC 2340  
AATACCGAGT TTGTCCTTAG GGTGAAATAT GATGTAGAGG GTCTTGAGCC AGGTCTCTAT 2400  
GTTGGAAGGA TAATCATTGA TGATCCAACA ACGCCAGTTA TTGAAGACGA GATCTTGAAC 2460  
ACAATTGTTA TTCCCGAGAA GTTCACTCCT GAGAACAATT ACACCCTCAC CTGGTATGAT 2520  
ATTAATGGTC CAGAAATGGT GACTCACCAC TTCTTCACTG TGCCTGAGGG AGTGGACGTT 2580  
CTCTACGCGA TGACCACATA CTGGGACTAC GGTCTGTACA GACCAGATGG AATGTTTGTG 2640  
TTCCCATACC AGCTAGATTA TCTTCCCGCT GCAGTCTCAA ATCCAATGCC TGGAACTGG 2700  
GAGCTAGTAT GGACTGGATT TAACTTTGCA CCCCTCTATG AGTCGGGCTT CCTTGTAAGG 2760  
ATTTACGGAG TAGAGATAAC TCCAAGCGTT TGGTACATTA ACAGGACATA CCTTGACACT 2820  
AACACTGAAT TCTCAATTGA ATTCAATATT ACTAACATCT ATGCCCCAAT TAATGCAACT 2880  
CTAATCCCCA TTGGCCTTGG AACCTACAAT GCGAGCGTTG AAAGCGTTGG TGATGGAGAG 2940  
TTCTTCATAA AGGGCATTGA AGTTCCTGAA GGCACCGCAG AGTTGAAGAT TAGGATAGGC 3000

03453-0440

AACCCAAGTG TTCCGAATTC AGATCTAGAC TTGTACCTTT ATGACAGTAA AGGCAATTTA 3060  
 GTGGCCTTAG ATGGAAACCC AACAGCAGAA GAAGAGGTTG TAGTTGAGTA TCCTAAGCCT 3120  
 GGAGTTTATT CAATAGTAGT ACATGGTTAC AGCGTCAGGG ACGAAAATGG TAATCCAACG 3180  
 ACAACCACCT TTGACTTAGT TGTTCAAATG ACCCTTGATA ATGGAAACAT AAAGCTTGAC 3240  
 AAAGACTCGA TTATTCTTGG AAGCAATGAA AGCGTAGTTG TAACTGCAAA CATAACAATT 3300  
 GATAGAGATC ATCCTACAGG AGTATACTCT GGTATCATAG AGATTAGAGA TAATGAGGTC 3360  
 TACCAGGATA CAAATACTTC AATTGCGAAA ATACCCATAA CTTTGGTAAT TGACAAGGCG 3420  
 GACTTTGCCG TTGGTCTCAC ACCAGCAGAG GGAGTACTTG GAGAGGCTAG AAATTACACT 3480  
 CTAATTGTAA AGCATGCCCT AACACTAGAG CCTGTGCCAA ATGCTACAGT GATTATAGGA 3540  
 AACTACACCT ACCTCACAGA CGAAAACGGT ACAGTGACAT TCACGTATGC TCCAACTAAG 3600  
 TTAGGCAGTG ATGAAATCAC AGTCATAGTT AAGAAAGAGA ACTTCAACAC ATTAGAGAAG 3660  
 ACCTTCCAAA TCACAGTATC AGAGCCTGAA ATAACTGAAG AGGACATAAA TGAGCCCAAG 3720  
 CTTGCAATGT CATCACCAGA AGCAAATGCT ACCATAGTAT CAGTTGAGAT GGAGAGTGAG 3780  
 GGTGGCGTTA AAAAGACAGT GACAGTGGA ATAACTATAA ACGGAACCGC TAATGAGACT 3840  
 GCAACAATAG TGGTTCCTGT TCCTAAGAAG GCCGAAAACA TCGAGGTAAG TGGAGACCAC 3900  
 GTAATTTCTT ATAGTATAGA GGAAGGAGAG TACGCCAAGT ACGTTATAAT TACAGTGAAG 3960  
 TTTGCATCAC CTGTAACAGT AACTGTTACT TACACTATCT ATGCTGGCCC AAGAGTCTCA 4020  
 ATCTTGACAC TTAAGTTCCT TGGCTACTCA TGGTACAGAC TATATTCACA GAAGTTTGAC 4080  
 GAATTGTACC AAAAGGCCCT TGAATTGGGA GTGGACAACG AGACATTAGC TTTAGCCCTC 4140  
 AGCTACCATG AAAAAGCCAA AGAGTACTAC GAAAAGGCCC TTGAGCTTAG CGAGGGTAAC 4200  
 ATAATCCAAT ACCTTGGAGA CATAAGACTA TTACCTCCAT TAAGACAGGC ATACATCAAT 4260  
 GAAATGAAGG CAGTTAAGAT ACTGGAAAAG GCCATAGAAG AATTAGAGGG TGAAGAGTAA 4320  
 TCTCCAATTT TTCCCACTTT TTCTTTTATA ACATTCCAAG CCTTTTCTTA GCTTCTTCGC 4380  
 TCATTCTATC AGGAGTCCAT GGAGGATCAA AGGTAAGTTC AACCTCCACA TCTCTTACTC 4440  
 CTGGGATTTT GAGTACTTTC TCCTCTACAG CTCTAAGAAG CCAGAGAGTT AAAGGACACC 4500  
 CAGGAGTTGT CATTGTCATC TTTATATATA CCGTTTTGTC AGGATTAATC TTTAGCTCAT 4560  
 AAATTAATCC AAGGTTTACA ACATCCATCC CAATTTCTGG GTCGATAACC TCCTTTAGCT 4620  
 TTTCCAGAAT CATTTCTTCA GTAATTTCAA GGTTCTCATC TTTGGTTTCT CTCACAAACC 4680  
 CAATTTCAAC CTGCCTGATA CCTTCTAACT CCCTAAGCTT GTTATATATC TCCAAAAGAG 4740  
 TGGCATCATC AATTTTCTCT TTAAG 4765

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1398 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Asn	Lys	Lys	Gly	Leu	Thr	Val	Leu	Phe	Ile	Ala	Ile	Met	Leu	5	10	15
Leu	Ser	Val	Val	Pro	Val	His	Phe	Val	Ser	Ala	Glu	Thr	Pro	Pro	20	25	30
Val	Ser	Ser	Glu	Asn	Ser	Thr	Thr	Ser	Ile	Leu	Pro	Asn	Gln	Gln	35	40	45
Val	Val	Thr	Lys	Glu	Val	Ser	Gln	Ala	Ala	Leu	Asn	Ala	Ile	Met	50	55	60
Lys	Gly	Gln	Pro	Asn	Met	Val	Leu	Ile	Ile	Lys	Thr	Lys	Glu	Gly	65	70	75
Lys	Leu	Glu	Glu	Ala	Lys	Thr	Glu	Leu	Glu	Lys	Leu	Gly	Ala	Glu	80	85	90
Ile	Leu	Asp	Glu	Asn	Arg	Val	Leu	Asn	Met	Leu	Leu	Val	Lys	Ile	95	100	105
Lys	Pro	Glu	Lys	Val	Lys	Glu	Leu	Asn	Tyr	Ile	Ser	Ser	Leu	Glu	110	115	120
Lys	Ala	Trp	Leu	Asn	Arg	Glu	Val	Lys	Leu	Ser	Pro	Pro	Ile	Val	125	130	135
Glu	Lys	Asp	Val	Lys	Thr	Lys	Glu	Pro	Ser	Leu	Glu	Pro	Lys	Met	140	145	150
Tyr	Asn	Ser	Thr	Trp	Val	Ile	Asn	Ala	Leu	Gln	Phe	Ile	Gln	Glu	155	160	165
Phe	Gly	Tyr	Asp	Gly	Ser	Gly	Val	Val	Val	Ala	Val	Leu	Asp	Thr	170	175	180
Gly	Val	Asp	Pro	Asn	His	Pro	Phe	Leu	Ser	Ile	Thr	Pro	Asp	Gly	185	190	195
Arg	Arg	Lys	Ile	Ile	Glu	Trp	Lys	Asp	Phe	Thr	Asp	Glu	Gly	Phe	200	205	210
Val	Asp	Thr	Ser	Phe	Ser	Phe	Ser	Lys	Val	Val	Asn	Gly	Thr	Leu	215	220	225
Ile	Ile	Asn	Thr	Thr	Phe	Gln	Val	Ala	Ser	Gly	Leu	Thr	Leu	Asn	230	235	240
Glu	Ser	Thr	Gly	Leu	Met	Glu	Tyr	Val	Val	Lys	Thr	Val	Tyr	Val	245	250	255
Ser	Asn	Val	Thr	Ile	Gly	Asn	Ile	Thr	Ser	Ala	Asn	Gly	Ile	Tyr	260	265	270
His	Phe	Gly	Leu	Leu	Pro	Glu	Arg	Tyr	Phe	Asp	Leu	Asn	Phe	Asp	275	280	285
Gly	Asp	Gln	Glu	Asp	Phe	Tyr	Pro	Val	Leu	Leu	Val	Asn	Ser	Thr	290	295	300

004453.04401

Gly	Asn	Gly	Tyr	Asp	Ile	Ala	Tyr	Val	Asp	Thr	Asp	Leu	Asp	Tyr	305	310	315
Asp	Phe	Thr	Asp	Glu	Val	Pro	Leu	Gly	Gln	Tyr	Asn	Val	Thr	Tyr	320	325	330
Asp	Val	Ala	Val	Phe	Ser	Tyr	Tyr	Tyr	Gly	Pro	Leu	Asn	Tyr	Val	335	340	345
Leu	Ala	Glu	Ile	Asp	Pro	Asn	Gly	Glu	Tyr	Ala	Val	Phe	Gly	Trp	350	355	360
Asp	Gly	His	Gly	His	Gly	Thr	His	Val	Ala	Gly	Thr	Val	Ala	Gly	365	370	375
Tyr	Asp	Ser	Asn	Asn	Asp	Ala	Trp	Asp	Trp	Leu	Ser	Met	Tyr	Ser	380	385	390
Gly	Glu	Trp	Glu	Val	Phe	Ser	Arg	Leu	Tyr	Gly	Trp	Asp	Tyr	Thr	395	400	405
Asn	Val	Thr	Thr	Asp	Thr	Val	Gln	Gly	Val	Ala	Pro	Gly	Ala	Gln	410	415	420
Ile	Met	Ala	Ile	Arg	Val	Leu	Arg	Ser	Asp	Gly	Arg	Gly	Ser	Met	425	430	435
Trp	Asp	Ile	Ile	Glu	Gly	Met	Thr	Tyr	Ala	Ala	Thr	His	Gly	Ala	440	445	450
Asp	Val	Ile	Ser	Met	Ser	Leu	Gly	Gly	Asn	Ala	Pro	Tyr	Leu	Asp	455	460	465
Gly	Thr	Asp	Pro	Glu	Ser	Val	Ala	Val	Asp	Glu	Leu	Thr	Glu	Lys	470	475	480
Tyr	Gly	Val	Val	Phe	Val	Ile	Ala	Ala	Gly	Asn	Glu	Gly	Pro	Gly	485	490	495
Ile	Asn	Ile	Val	Gly	Ser	Pro	Gly	Val	Ala	Thr	Lys	Ala	Ile	Thr	500	505	510
Val	Gly	Ala	Ala	Ala	Val	Pro	Ile	Asn	Val	Gly	Val	Tyr	Val	Ser	515	520	525
Gln	Ala	Leu	Gly	Tyr	Pro	Asp	Tyr	Tyr	Gly	Phe	Tyr	Tyr	Phe	Pro	530	535	540
Ala	Tyr	Thr	Asn	Val	Arg	Ile	Ala	Phe	Phe	Ser	Ser	Arg	Gly	Pro	545	550	555
Arg	Ile	Asp	Gly	Glu	Ile	Lys	Pro	Asn	Val	Val	Ala	Pro	Gly	Tyr	560	565	570
Gly	Ile	Tyr	Ser	Ser	Leu	Pro	Met	Trp	Ile	Gly	Gly	Ala	Asp	Phe	575	580	585
Met	Ser	Gly	Thr	Ser	Met	Ala	Thr	Pro	His	Val	Ser	Gly	Val	Val	590	595	600
Ala	Leu	Leu	Ile	Ser	Gly	Ala	Lys	Ala	Glu	Gly	Ile	Tyr	Tyr	Asn	605	610	615
Pro	Asp	Ile	Ile	Lys	Lys	Val	Leu	Glu	Ser	Gly	Ala	Thr	Trp	Leu	620	625	630

034155-0444

Glu	Gly	Asp	Pro	Tyr	Thr	Gly	Gln	Lys	Tyr	Thr	Glu	Leu	Asp	Gln
				635					640					645
Gly	His	Gly	Leu	Val	Asn	Val	Thr	Lys	Ser	Trp	Glu	Ile	Leu	Lys
				650					655					660
Ala	Ile	Asn	Gly	Thr	Thr	Leu	Pro	Ile	Val	Asp	His	Trp	Ala	Asp
				665					670					675
Lys	Ser	Tyr	Ser	Asp	Phe	Ala	Glu	Tyr	Leu	Gly	Val	Asp	Val	Ile
				680					685					690
Arg	Gly	Leu	Tyr	Ala	Arg	Asn	Ser	Ile	Pro	Asp	Ile	Val	Glu	Trp
				695					700					705
His	Ile	Lys	Tyr	Val	Gly	Asp	Thr	Glu	Tyr	Arg	Thr	Phe	Glu	Ile
				710					715					720
Tyr	Ala	Thr	Glu	Pro	Trp	Ile	Lys	Pro	Phe	Val	Ser	Gly	Ser	Val
				725					730					735
Ile	Leu	Glu	Asn	Asn	Thr	Glu	Phe	Val	Leu	Arg	Val	Lys	Tyr	Asp
				740					745					750
Val	Glu	Gly	Leu	Glu	Pro	Gly	Leu	Tyr	Val	Gly	Arg	Ile	Ile	Ile
				755					760					765
Asp	Asp	Pro	Thr	Thr	Pro	Val	Ile	Glu	Asp	Glu	Ile	Leu	Asn	Thr
				770					775					780
Ile	Val	Ile	Pro	Glu	Lys	Phe	Thr	Pro	Glu	Asn	Asn	Tyr	Thr	Leu
				785					790					795
Thr	Trp	Tyr	Asp	Ile	Asn	Gly	Pro	Glu	Met	Val	Thr	His	His	Phe
				800					805					810
Phe	Thr	Val	Pro	Glu	Gly	Val	Asp	Val	Leu	Tyr	Ala	Met	Thr	Thr
				815					820					825
Tyr	Trp	Asp	Tyr	Gly	Leu	Tyr	Arg	Pro	Asp	Gly	Met	Phe	Val	Phe
				830					835					840
Pro	Tyr	Gln	Leu	Asp	Tyr	Leu	Pro	Ala	Ala	Val	Ser	Asn	Pro	Met
				845					850					855
Pro	Gly	Asn	Trp	Glu	Leu	Val	Trp	Thr	Gly	Phe	Asn	Phe	Ala	Pro
				860					865					870
Leu	Tyr	Glu	Ser	Gly	Phe	Leu	Val	Arg	Ile	Tyr	Gly	Val	Glu	Ile
				875					880					885
Thr	Pro	Ser	Val	Trp	Tyr	Ile	Asn	Arg	Thr	Tyr	Leu	Asp	Thr	Asn
				890					895					900
Thr	Glu	Phe	Ser	Ile	Glu	Phe	Asn	Ile	Thr	Asn	Ile	Tyr	Ala	Pro
				905					910					915
Ile	Asn	Ala	Thr	Leu	Ile	Pro	Ile	Gly	Leu	Gly	Thr	Tyr	Asn	Ala
				920					925					930
Ser	Val	Glu	Ser	Val	Gly	Asp	Gly	Glu	Phe	Phe	Ile	Lys	Gly	Ile
				935					940					945
Glu	Val	Pro	Glu	Gly	Thr	Ala	Glu	Leu	Lys	Ile	Arg	Ile	Gly	Asn
				950					955					960



Pro Ser Val	Pro Asn	Ser Asp	Leu Asp	Leu Tyr	Leu Tyr	Asp Ser	
	965			970		975	
Lys Gly Asn	Leu Val	Ala Leu	Asp Gly	Asn Pro	Thr Ala	Glu Glu	
	980			985		990	
Glu Val Val	Val Glu	Tyr Pro	Lys Pro	Gly Val	Tyr Ser	Ile Val	
	995			1000		1005	
Val His Gly	Tyr Ser	Val Arg	Asp Glu	Asn Gly	Asn Pro	Thr Thr	
	1010			1015		1020	
Thr Thr Phe	Asp Leu	Val Val	Gln Met	Thr Leu	Asp Asn	Gly Asn	
	1025			1030		1035	
Ile Lys Leu	Asp Lys	Asp Ser	Ile Ile	Leu Gly	Ser Asn	Glu Ser	
	1040			1045		1050	
Val Val Val	Thr Ala	Asn Ile	Thr Ile	Asp Arg	Asp His	Pro Thr	
	1055			1060		1065	
Gly Val Tyr	Ser Gly	Ile Ile	Glu Ile	Arg Asp	Asn Glu	Val Tyr	
	1070			1075		1080	
Gln Asp Thr	Asn Thr	Ser Ile	Ala Lys	Ile Pro	Ile Thr	Leu Val	
	1085			1090		1095	
Ile Asp Lys	Ala Asp	Phe Ala	Val Gly	Leu Thr	Pro Ala	Glu Gly	
	1100			1105		1110	
Val Leu Gly	Glu Ala	Arg Asn	Tyr Thr	Leu Ile	Val Lys	His Ala	
	1115			1120		1125	
Leu Thr Leu	Glu Pro	Val Pro	Asn Ala	Thr Val	Ile Ile	Gly Asn	
	1130			1135		1140	
Tyr Thr Tyr	Leu Thr	Asp Glu	Asn Gly	Thr Val	Thr Phe	Thr Tyr	
	1145			1150		1155	
Ala Pro Thr	Lys Leu	Gly Ser	Asp Glu	Ile Thr	Val Ile	Val Lys	
	1160			1165		1170	
Lys Glu Asn	Phe Asn	Thr Leu	Glu Lys	Thr Phe	Gln Ile	Thr Val	
	1175			1180		1185	
Ser Glu Pro	Glu Ile	Thr Glu	Glu Asp	Ile Asn	Glu Pro	Lys Leu	
	1190			1195		1200	
Ala Met Ser	Ser Pro	Glu Ala	Asn Ala	Thr Ile	Val Ser	Val Glu	
	1205			1210		1215	
Met Glu Ser	Glu Gly	Gly Val	Lys Lys	Thr Val	Thr Val	Glu Ile	
	1220			1225		1230	
Thr Ile Asn	Gly Thr	Ala Asn	Glu Thr	Ala Thr	Ile Val	Val Pro	
	1235			1240		1245	
Val Pro Lys	Lys Ala	Glu Asn	Ile Glu	Val Ser	Gly Asp	His Val	
	1250			1255		1260	
Ile Ser Tyr	Ser Ile	Glu Glu	Gly Glu	Tyr Ala	Lys Tyr	Val Ile	
	1265			1270		1275	
Ile Thr Val	Lys Phe	Ala Ser	Pro Val	Thr Val	Thr Val	Thr Tyr	
	1280			1285		1290	

094553 042401

Thr Ile Tyr Ala Gly Pro Arg Val Ser Ile Leu Thr Leu Asn Phe  
1295 1300 1305

Leu Gly Tyr Ser Trp Tyr Arg Leu Tyr Ser Gln Lys Phe Asp Glu  
1310 1315 1320

Leu Tyr Gln Lys Ala Leu Glu Leu Gly Val Asp Asn Glu Thr Leu  
1325 1330 1335

Ala Leu Ala Leu Ser Tyr His Glu Lys Ala Lys Glu Tyr Tyr Glu  
1340 1345 1350

Lys Ala Leu Glu Leu Ser Glu Gly Asn Ile Ile Gln Tyr Leu Gly  
1355 1360 1365

Asp Ile Arg Leu Leu Pro Pro Leu Arg Gln Ala Tyr Ile Asn Glu  
1370 1375 1380

Met Lys Ala Val Lys Ile Leu Glu Lys Ala Ile Glu Glu Leu Glu  
1385 1390 1395

Gly Glu Glu

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGWWSDRRTG TTRRHGTHGC DGTDMTYGAC ACBGG

35

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

KSTCACGGAA CTCACGTDGC BGGMACDGTT GC

32

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ASCMGCAACH GTKCCVGCHA CGTGAGTTCC GTG

33

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 34 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CHCCGSYVAC RTGBGGAGWD GCCATBGAVG TDCC

34

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 145 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

A GTT GCG GTA ATT GAC ACG GGT ATA GAC GCG AAC CAC CCC GAT CTG	46
Val Ala Val Ile Asp Thr Gly Ile Asp Ala Asn His Pro Asp Leu	
5 10 15	
AAG GGC AAG GTC ATA GGC TGG TAC GAC GCC GTC AAC GGC AGG TCG	91
Lys Gly Lys Val Ile Gly Trp Tyr Asp Ala Val Asn Gly Arg Ser	
20 25 30	
ACC CCC TAC GAT GAC CAG GGA CAC GGA ACT CAC GTN GCN GGA ACN	136
Thr Pro Tyr Asp Asp Gln Gly His Gly Thr His Val Ala Gly Thr	
35 40 45	
GTT GCT GGT	145
Val Ala Gly	

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 564 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCT CAC GGA ACT CAC GTG GCG GGA ACA GTT GCC GGA ACA GGC AGC	45
Ser His Gly Thr His Val Ala Gly Thr Val Ala Gly Thr Gly Ser	
5 10 15	
GTT AAC TCC CAG TAC ATA GGC GTC GCC CCC GGC GCG AAG CTC GTC	90
Val Asn Ser Gln Tyr Ile Gly Val Ala Pro Gly Ala Lys Leu Val	
20 25 30	
GGT GTC AAG GTT CTC GGT GCC GAC GGT TCG GGA AGC GTC TCC ACC	135
Gly Val Lys Val Leu Gly Ala Asp Gly Ser Gly Ser Val Ser Thr	
35 40 45	

ATC	ATC	GCG	GGT	GTT	GAC	TGG	GTC	GTC	CAG	AAC	AAG	GAT	AAG	TAC	180
Ile	Ile	Ala	Gly	Val	Asp	Trp	Val	Val	Gln	Asn	Lys	Asp	Lys	Tyr	
				50					55					60	
GGG	ATA	AGG	GTC	ATC	AAC	CTC	TCC	CTC	GGC	TCC	TCC	CAG	AGC	TCC	225
Gly	Ile	Arg	Val	Ile	Asn	Leu	Ser	Leu	Gly	Ser	Ser	Gln	Ser	Ser	
				65					70					75	
GAC	GGA	GCC	GAC	TCC	CTC	AGT	CAG	GCC	GTC	AAC	AAC	GCC	TGG	GAC	270
Asp	Gly	Ala	Asp	Ser	Leu	Ser	Gln	Ala	Val	Asn	Asn	Ala	Trp	Asp	
				80					85					90	
GCC	GGT	ATA	GTA	GTC	TGC	GTC	GCC	GCC	GGC	AAC	AGC	GGG	CCG	AAC	315
Ala	Gly	Ile	Val	Val	Cys	Val	Ala	Ala	Gly	Asn	Ser	Gly	Pro	Asn	
				95					100					105	
ACC	TAC	ACC	GTC	GGC	TCA	CCC	GCC	GCC	GCG	AGC	AAG	GTC	ATA	ACC	360
Thr	Tyr	Thr	Val	Gly	Ser	Pro	Ala	Ala	Ala	Ser	Lys	Val	Ile	Thr	
				110					115					120	
GTC	GGT	GCA	GTT	GAC	AGC	AAC	GAC	AAC	ATC	GCC	AGC	TTC	TCC	AGC	405
Val	Gly	Ala	Val	Asp	Ser	Asn	Asp	Asn	Ile	Ala	Ser	Phe	Ser	Ser	
				125					130					135	
AGG	GGA	CCG	ACC	GCG	GAC	GGA	AGG	CTC	AAG	CCG	GAA	GTC	GTC	GCC	450
Arg	Gly	Pro	Thr	Ala	Asp	Gly	Arg	Leu	Lys	Pro	Glu	Val	Val	Ala	
				140					145					150	
CCC	GGC	GTT	GAC	ATC	ATA	GCC	CCG	CGC	GCC	AGC	GGA	ACC	AGC	ATG	495
Pro	Gly	Val	Asp	Ile	Ile	Ala	Pro	Arg	Ala	Ser	Gly	Thr	Ser	Met	
				155					160					165	
GGC	ACC	CCG	ATA	AAC	GAC	TAC	TAC	ACC	AAG	GCC	TCT	GGA	ACC	TCA	540
Gly	Thr	Pro	Ile	Asn	Asp	Tyr	Tyr	Thr	Lys	Ala	Ser	Gly	Thr	Ser	
				170					175					180	
ATG	GCC	ACT	CCC	CAT	GTT	ACC	GGT								564
Met	Ala	Thr	Pro	His	Val	Thr	Gly								
				185											

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1859 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAGCTCCGAC	GGAACCGACT	CCCTCAGTCA	GGCCGTCAAC	AACGCCTGGG	ACGCCGGTAT	60
AGTAGTCTGC	GTCGCCGCCG	GCAACAGCGG	GCCGAACACC	TACACCGTCG	GCTCACCCGC	120
CGCCGCGAGC	AAGGTCATAA	CCGTCGGTGC	AGTTGACAGC	AACGACAACA	TCGCCAGCTT	180
CTCCAGCAGG	GGACCGACCG	CGGACGGAAG	GCTCAAGCCG	GAAGTCGTCG	CCCCCGGCGT	240
TGACATCATA	GCCCCGCGCG	CCAGCGGAAC	CAGCATGGGC	ACCCCGATAA	ACGACTACTA	300
CACCAAGGCC	TCTGGAACCA	GCATGGCCAC	CCCGCACGTT	TCGGGCGTTG	GCGCGCTCAT	360
CCTCCAGGCC	CACCCGAGCT	GGACCCCGGA	CAAGGTGAAG	ACCGCCCTCA	TCGAGACCGC	420

CGACATAGTC GCCCCAAGG AGATAGCGGA CATCGCCTAC GGTGCGGGTA GGGTGAACGT 480  
 CTACAAGGCC ATCAAGTACG ACGACTACGC CAAGCTCACC TTCACCGGCT CCGTCGCCGA 540  
 CAAGGGAAGC GCCACCCACA CCTTCGACGT CAGCGGCGCC ACCTTCGTGA CCGCCACCCT 600  
 CTACTGGGAC ACGGGCTCGA GCGACATCGA CCTCTACCTC TACGACCCCA ACGGGAACGA 660  
 GGTTGACTAC TCCTACACCG CCTACTACGG CTTCGAGAAG GTCGGCTACT ACAACCCGAC 720  
 CGCCGGAACC TGGACGGTCA AGGTCGTCAG CTACAAGGGC GCGGCGAACT ACCAGGTCGA 780  
 CGTCGTCAGC GACGGGAGCC TCAGCCAGTC CGGCGGCGGC AACCCGAATC CAAACCCCAA 840  
 CCCGAACCCA ACCCCGACCA CCGACACCCA GACCTTCACC GGTCCGTTA ACGACTACTG 900  
 GGACACCAGC GACACCTTCA CCATGAACGT CAACAGCGGT GCCACCAAGA TAACCGGTGA 960  
 CCTGACCTTC GATACTTCCT ACAACGACCT CGACCTCTAC CTCTACGACC CCAACGGCAA 1020  
 CCTCGTTGAC AGGTCCACGT CGAGCAACAG CTACGAGCAC GTCGAGTACG CCAACCCCGC 1080  
 CCCGGGAACC TGGACGTTCC TCGTCTACGC CTACAGCACC TACGGCTGGG CGGACTACCA 1140  
 GCTCAAGGCC GTCGTCTACT ACGGGTGAAG GTTTTTAATC CCCTTTTCTT TCCCCTTTTG 1200  
 AGGTGGTTGG GATGAAGCGG GTTCTTGCGG CGATCCTTGT AATCATGCTC ATCGGATTAT 1260  
 CATTCCCTGC CGGAAGTGCT AAAATCGAGC CCTACGTTTA CAGCCCCACC GTTCCGGATA 1320  
 CCGCCTTCGC GGTTCCTACC CTGTACAGGA CCGGGGACTA CGCCCGGGTT CTCGAGGGAT 1380  
 ACGAGTGGCT CCTCCAGATG AGAACTCCCA TCGATTCTGT GGGGGTTTCC CGCGGGGAAA 1440  
 CGCACATGGC CAAGTACACG GCAATGGCGA TGCTGGCCCT CATGCGCGGC GAGAACGTGG 1500  
 CGAGGGGGCG TTACAGGGAT GTTCTCAACG ACGCCGCGTA CTGGTTAATA TACAAACAGA 1560  
 ACCCGGACGG CTCGTGGGAG GACTACACCG GAACGGCGCT GGCCGTCATC GCGCTCGGGG 1620  
 AGTTCCTTAA GGGCGGGTAC ATCAACGCGA ACCTGACCGG CTTCAAAAAG CAGGTTAAAG 1680  
 AGGCCGTAAA CCGCGGGGAA GGCTGGCTGA TGGATGCGGA CCCAAAACG GACGCGGATA 1740  
 GAATATTCGG CTACCTCGCC CTCGGTAAAA AGGACGAAC TAAAAAGATG AACCTTCCG 1800  
 GTGACCTGAA GGCCTACCGC GCCTTGCAC TTGCCTACCT CGGGGAGAGG GTCGAGCTC 1859

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TGTAAGTAGTC GTTTATCGGG

20

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1464 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AAGCTTAACA TCGAGCGCTC CACCTCTAAA GTAGGTGAGT GTGGATACGA AGGTTAGGGC	60
CGCTATGACG ACCTTCAGGA TCCCAACGGC TTCTTTTATG GGGAGCCCGG CGAAGGTGAG	120
AATTGAAAGG ATTACCATAC TCCCTCCGCT CATCATGGAG CCTATGAATC CCCCTCCAAA	180
AGAGAGAAGT GCTATAAGGA GCGTCCTCAT GTTCCATGCT ATGTTTTGGT ATTTAATGCT	240
TTTCCGCTTA ATGTTACACC TCCTCATGAC AATTCGCGT TTAGGGATGG GGTAAATTGG	300
ACCCCTCCGA GCCACGGGTT GATGTCCATT ATGTCGATAT TCACCATCTT ATCCCCAACT	360
TTGTGGGTTT CAAACATTAC CCTACGTTAT ATTTTATCG TCCTAATTAA CTGCTGAAAC	420
GGGCGCTTAT CGTTCATCGT TGATGGTTTT GGGTGACCGG GCATTAAGGA ATTGTGTCGT	480
TTGCTGAAAT TTATGAAACG GAGTTGGCTT CTTTATGTTA CATAAAGATG TACATTACTG	540
TAATGTATAT AAATGGAAGA AACACTGTTG CGTAAACTTT TTAATGTATC CAATATCAGT	600
ACTTCGATGT CCCGATATGG GACATGTTGG ATAGGAGGGT ACTGGAATGA AGAGGTTAGG	660
TGCTGTGGTG CTGGCACTGG TGCTCGTGGG TCTTCTGGCC GGAACGGCCC TTGCGGCACC	720
CGTAAAACCG GTTGTCAGGA ACAACGCGGT TCAGCAGAAG AACTACGGAC TGCTGACCCC	780
GGGACTGTTC AAGAAAGTCC AGAGGATGAA CTGGAACCAG GAAGTGGACA CCGTCATAAT	840
GTTCGGGAGC TACGGAGACA GGGACAGGGC GGTAAAGGTA CTGAGGCTCA TGGGCGCCCA	900
GGTCAAGTAC TCCTACAAGA TAATCCCTGC TGTCGCGGTT AAAATAAAGG CCAGGGACCT	960
TCTGCTGATC GCGGGCATGA TAGACACGGG TTA CTTCGGT AACACAAGGG TCTCGGGCAT	1020
AAAGTTCATA CAGGAGGATT ACAAGGTTCA GGTTGACGAC GCCACTTCCG TCTCCAGAT	1080
AGGGGCCGAT ACCGTCTGGA ACTCCCTCGG CTACGACGGA AGCGGTGTGG TG GTTGCCAT	1140
CGTCGATACG GGTATAGACG CGAACCACCC CGATCTGAAG GGCAAGGTCA TAGGCTGGTA	1200
CGACTCCGTC AACGGCAGGT CGACCCCTA CGATGACCAG GGACACGGAA CCCACGTTGC	1260
GGGTATCGTT GCCGGAACCG GGAGCGTTAA CTCCAGTAC ATAGGCGTCG GCGCGGCGC	1320
GAAGCTCGTC GCGTCAAGG TTCTCGGTTT CGACGGTTTC GGAAGCGTCT CCACCATCAT	1380
CGCGGGTGTT GACTGGAACG TCCAGAACTA GGACAAGTAC GGGATAAGGG TCATCAACCT	1440
CTCCCTCGGC TCCTCCAGG GCTC	1464

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AAAAGAATTC GGATCCATGA AGAGGTTAGG TGC

33

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTTTATCGAT CAGGCGTCCC AGGCGTTG

28

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CATTATAGGT AAGAGAGGAA TG

22

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GATCCATTCC TCTCTTACCT ATAATGGTAC

30

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TAGCAGTAAT TGACACGGG

19

TCTCTTACCT AAGAGAGGAA TG

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TAGCAGTAAT TGACACTGG

19

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CTGTTCCAGC TACGTGAGTT CC

22

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTGTTCCAGC TACATGAGTT CC

22

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 507 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

A	CTA	GTC	ATC	TCA	GGT	TTA	ACA	GGG	GGT	AAA	GCT	AAG	CTT	TCA	GGT	46
	Leu	Val	Ile	Ser	Gly	Leu	Thr	Gly	Gly	Lys	Ala	Lys	Leu	Ser	Gly	
					5					10					15	
	GTT	AGG	TTT	ATC	CAG	GAA	GAC	TAC	AAA	GTT	ACA	GTT	TCA	GCA	GAA	91
	Val	Arg	Phe	Ile	Gln	Glu	Asp	Tyr	Lys	Val	Thr	Val	Ser	Ala	Glu	
					20					25					30	
	TTA	GAA	GGA	CTG	GAT	GAG	TCT	GCA	GCT	CAA	GTT	ATG	GCA	ACT	TAC	136
	Leu	Glu	Gly	Leu	Asp	Glu	Ser	Ala	Ala	Gln	Val	Met	Ala	Thr	Tyr	
					35					40					45	



GTT TGG AAC TTG GGA TAT GAT GGT TCT GGA ATC ACA ATA GGA ATA	181
Val Trp Asn Leu Gly Tyr Asp Gly Ser Gly Ile Thr Ile Gly Ile	
50 55 60	
ATT GAC ACT GGA ATT GAC GCT TCT CAT CCA GAT CTC CAA GGA AAA	226
Ile Asp Thr Gly Ile Asp Ala Ser His Pro Asp Leu Gln Gly Lys	
65 70 75	
GTA ATT GGG TGG GTA GAT TTT GTC AAT GGT AGG AGT TAT CCA TAC	271
Val Ile Gly Trp Val Asp Phe Val Asn Gly Arg Ser Tyr Pro Tyr	
80 85 90	
GAT GAC CAT GGA CAT GGA ACT CAT GTA GCT TCA ATA GCA GCT GGT	316
Asp Asp His Gly His Gly Thr His Val Ala Ser Ile Ala Ala Gly	
95 100 105	
ACT GGA GCA GCA AGT AAT GGC AAG TAC AAG GGA ATG GCT CCA GGA	361
Thr Gly Ala Ala Ser Asn Gly Lys Tyr Lys Gly Met Ala Pro Gly	
110 115 120	
GCT AAG CTG GCG GGA ATT AAG GTT CTA GGT GCC GAT GGT TCT GGA	406
Ala Lys Leu Ala Gly Ile Lys Val Leu Gly Ala Asp Gly Ser Gly	
125 130 135	
AGC ATA TCT ACT ATA ATT AAG GGA GTT GAG TGG GCC GTT GAT AAC	451
Ser Ile Ser Thr Ile Ile Lys Gly Val Glu Trp Ala Val Asp Asn	
140 145 150	
AAA GAT AAG TAC GGA ATT AAG GTC ATT AAT CTT TCT CTT GGT TCA	496
Lys Asp Lys Tyr Gly Ile Lys Val Ile Asn Leu Ser Leu Gly Ser	
155 160 165	
AGC CAG AGC TC	507
Ser Gln Ser	

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGACACTGGA ATTGACGCTT CTCATCCAGA 30

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TCTCCAAGGA AAAGTAATTG GGTGGGTAGA 30

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 30 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GTGCGCATAA CTTGAGCTGC AGACTCATCC

30

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 419 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTTATTAAGC ATAAAATAGC CATGCAACTT TGATCACTAA TGTGCGGTGG TGCAC ATG	58
	Met
AAG GGG CTG AAA GCT CTC ATA TTA GTG ATT TTA GTT CTA GGT TTG	103
Lys Gly Leu Lys Ala Leu Ile Leu Val Ile Leu Val Leu Gly Leu	
5 10 15	
GTA GTA GGG AGC GTA GCG GCA GCT CCA GAG AAG AAA GTT GTT CAA	148
Val Val Gly Ser Val Ala Ala Ala Pro Glu Lys Lys Val Val Gln	
20 25 30	
GTA AGA AAT GTT GAG AAG AAC TAT GGT CTG CTA ACG CCA GGA CTG	193
Val Arg Asn Val Glu Lys Asn Tyr Gly Leu Leu Thr Pro Gly Leu	
35 40 45	
TTC AGA AAA ATT CCC AAA TTG GAT CCT AAC GAG GGA ATC AGC ACA	238
Phe Arg Lys Ile Pro Lys Leu Asp Pro Asn Glu Gly Ile Ser Thr	
50 55 60	
GTA ATT GTA TTT GTT AAC CAT AGG GGA AAA GAA ATT GCA GTA AGA	283
Val Ile Val Phe Val Asn His Arg Gly Lys Glu Ile Ala Val Arg	
65 70 75	
GTT CTT GAG TTA ATG GGT GCC CAA GTT AGG TAT GTG TAC CAT ATT	328
Val Leu Glu Leu Met Gly Ala Gln Val Arg Tyr Val Tyr His Ile	
80 85 90	
ATA CCC CCA ATA GCT GCC GAT CTT AAG GTT AGA GAC TTA CTA GTC	373
Ile Pro Pro Ile Ala Ala Asp Leu Lys Val Arg Asp Leu Leu Val	
95 100 105	
ATC TCA GGT TTA ACA GGG GGT GAA ACT AAG CTT TCA GGT GTT AGG T	419
Ile Ser Gly Leu Thr Gly Gly Glu Thr Lys Leu Ser Gly Val Arg	
110 115 120	

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 180 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double



AGGTTTATCC AGGAAGACTA CAAAGTTACA GTTTCAGCAG AATTAGAAGG ACTGGATGAG 420  
TCTGCAGCTC AAGTTATGGC AACTTACGTT TGGAACTTGG GATATGATGG TTCTGGAATC 480  
ACAATAGGAA TAATTGACAC TGGAAATTGAC GCTTCTCATC CAGATCTCCA AGGAAAAGTA 540  
ATTGGGTGGG TAGATTTTGT CAATGGTAGG AGTTATCCAT ACGATGACCA TGGACATGGA 600  
ACTCATGTAG CTTCAATAGC AGCTGGTACT GGAGCAGCAA GTAATGGCAA GTACAAGGGA 660  
ATGGCTCCAG GAGCTAAGCT GGCGGGAATT AAGGTTCTAG GTGCCGATGG TTCTGGAAGC 720  
ATATCTACTA TAATTAAGGG AGTTGAGTGG GCCGTTGATA ACAAAGATAA GTACGGAATT 780  
AAGGTCATTA ATCTTCTCT TGGTTCAAGC CAGAGCTCAG ATGGTACTGA CGCTCTAAGT 840  
CAGGCTGTTA ATGCAGCGTG GGATGCTGGA TTAGTTGTTG TGGTTGCCGC TGGAAACAGT 900  
GGACCTAACA AGTATACAAT CGGTTCTCCA GCAGCTGCAA GCAAAGTTAT TACAGTTGGA 960  
GCCGTTGACA AGTATGATGT TATAACAAGC TTCTCAAGCA GAGGGCCAAC TGCAGACGGC 1020  
AGGCTTAAGC CTGAGGTTGT TGCTCCAGGA AACTGGATAA TTGCTGCCAG AGCAAGTGGA 1080  
ACTAGCATGG GTCAACCAAT TAATGACTAT TACACAGCAG CTCCTGGGAC ATCAATGGCA 1140  
ACTCCTCAG TAGCTGGTAT TGCAGCCCTC TTGCTCCAAG CACACCCGAG CTGGACTCCA 1200  
GACAAAGTAA AAACAGCCCT CATAGAAACT GCTGATATCG TAAAGCCAGA TGAAATAGCC 1260  
GATATAGCCT ACGGTGCAGG TAGGGTTAAT GCATACAAGG CTATAAACTA CGATAACTAT 1320  
GCAAAGCTAG TGTTCACTGG ATATGTTGCC AACAAAGGCA GCCAAACTCA CCAGTTCGTT 1380  
ATTAGCGGAG CTTGTTTCGT AACTGCCACA TTATACTGGG ACAATGCCAA TAGCGACCTT 1440  
GATCTTTACC TCTACGATCC CAATGGAAAC CAGGTTGACT ACTCTTACAC CGCCTACTAT 1500  
GGATTGAAA AGGTTGGTTA TTACAACCCA ACTGATGGAA CATGGACAAT TAAGGTTGTA 1560  
AGCTACAGCG GAAGTGCAA CTATCAAGTA GATGTGGTAA GTGATGGTTC CCTTTCACAG 1620  
CCTGGAAGTT CACCATCTCC ACAACCAGAA CCAACAGTAG ACGCAAAGAC GTTCCAAGGA 1680  
TCCGATCACT ACTACTATGA CAGGAGCGAC ACCTTTACAA TGACCGTTAA CTCTGGGGCT 1740  
ACAAAGATTA CTGGAGACCT AGTGTTTGAC ACAAGCTACC ATGATCTTGA CCTTTACCTC 1800  
TACGATCCTA ACCAGAAGCT TGTAGATAGA TCGGAGAGTC CCAACAGCTA CGAACACGTA 1860  
GAATACTTAA CCCCCGCCCC AGGAACCTGG TACTTCCTAG TATATGCCTA CTACACTTAC 1920  
GGTTGGGCTT ACTACGAGCT GACGGCTAAA GTTTATTATG GC 1962

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 654 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met	Lys	Gly	Leu	Lys	Ala	Leu	Ile	Leu	Val	Ile	Leu	Val	Leu	Gly	5	10	15
Leu	Val	Val	Gly	Ser	Val	Ala	Ala	Ala	Pro	Glu	Lys	Lys	Val	Glu	20	25	30
Gln	Val	Arg	Asn	Val	Glu	Lys	Asn	Tyr	Gly	Leu	Leu	Thr	Pro	Gly	35	40	45
Leu	Phe	Arg	Lys	Ile	Gln	Lys	Leu	Asn	Pro	Asn	Glu	Glu	Ile	Ser	50	55	60
Thr	Val	Ile	Val	Phe	Glu	Asn	His	Arg	Glu	Lys	Glu	Ile	Ala	Val	65	70	75
Arg	Val	Leu	Glu	Leu	Met	Gly	Ala	Lys	Val	Arg	Tyr	Val	Tyr	His	80	85	90
Ile	Ile	Pro	Ala	Ile	Ala	Ala	Asp	Leu	Lys	Val	Arg	Asp	Leu	Leu	95	100	105
Val	Ile	Ser	Gly	Leu	Thr	Gly	Gly	Lys	Ala	Lys	Leu	Ser	Gly	Val	110	115	120
Arg	Phe	Ile	Gln	Glu	Asp	Tyr	Lys	Val	Thr	Val	Ser	Ala	Glu	Leu	125	130	135
Glu	Gly	Leu	Asp	Glu	Ser	Ala	Ala	Gln	Val	Met	Ala	Thr	Tyr	Val	140	145	150
Trp	Asn	Leu	Gly	Tyr	Asp	Gly	Ser	Gly	Ile	Thr	Ile	Gly	Ile	Ile	155	160	165
Asp	Thr	Gly	Ile	Asp	Ala	Ser	His	Pro	Asp	Leu	Gln	Gly	Lys	Val	170	175	180
Ile	Gly	Trp	Val	Asp	Phe	Val	Asn	Gly	Arg	Ser	Tyr	Pro	Tyr	Asp	185	190	195
Asp	His	Gly	His	Gly	Thr	His	Val	Ala	Ser	Ile	Ala	Ala	Gly	Thr	200	205	210
Gly	Ala	Ala	Ser	Asn	Gly	Lys	Tyr	Lys	Gly	Met	Ala	Pro	Gly	Ala	215	220	225
Lys	Leu	Ala	Gly	Ile	Lys	Val	Leu	Gly	Ala	Asp	Gly	Ser	Gly	Ser	230	235	240
Ile	Ser	Thr	Ile	Ile	Lys	Gly	Val	Glu	Trp	Ala	Val	Asp	Asn	Lys	245	250	255
Asp	Lys	Tyr	Gly	Ile	Lys	Val	Ile	Asn	Leu	Ser	Leu	Gly	Ser	Ser	260	265	270
Gln	Ser	Ser	Asp	Gly	Thr	Asp	Ala	Leu	Ser	Gln	Ala	Val	Asn	Ala	275	280	285
Ala	Trp	Asp	Ala	Gly	Leu	Val	Val	Val	Val	Ala	Ala	Gly	Asn	Ser	290	295	300
Gly	Pro	Asn	Lys	Tyr	Thr	Ile	Gly	Ser	Pro	Ala	Ala	Ala	Ser	Lys	305	310	315
Val	Ile	Thr	Val	Gly	Ala	Val	Asp	Lys	Tyr	Asp	Val	Ile	Thr	Ser	320	325	330

0945-0440

Phe	Ser	Ser	Arg	Gly 335	Pro	Thr	Ala	Asp	Gly 340	Arg	Leu	Lys	Pro	Glu 345
Val	Val	Ala	Pro	Gly 350	Asn	Trp	Ile	Ile	Ala 355	Ala	Arg	Ala	Ser	Gly 360
Thr	Ser	Met	Gly	Gln 365	Pro	Ile	Asn	Asp	Tyr 370	Tyr	Thr	Ala	Ala	Pro 375
Gly	Thr	Ser	Met	Ala 380	Thr	Pro	His	Val	Ala 385	Gly	Ile	Ala	Ala	Leu 390
Leu	Leu	Gln	Ala	His 395	Pro	Ser	Trp	Thr	Pro 400	Asp	Lys	Val	Lys	Thr 405
Ala	Leu	Ile	Glu	Thr 410	Ala	Asp	Ile	Val	Lys 415	Pro	Asp	Glu	Ile	Ala 420
Asp	Ile	Ala	Tyr	Gly 425	Ala	Gly	Arg	Val	Asn 430	Ala	Tyr	Lys	Ala	Ile 435
Asn	Tyr	Asp	Asn	Tyr 440	Ala	Lys	Leu	Val	Phe 445	Thr	Gly	Tyr	Val	Ala 450
Asn	Lys	Gly	Ser	Gln 455	Thr	His	Gln	Phe	Val 460	Ile	Ser	Gly	Ala	Ser 465
Phe	Val	Thr	Ala	Thr 470	Leu	Tyr	Trp	Asp	Asn 475	Ala	Asn	Ser	Asp	Leu 480
Asp	Leu	Tyr	Leu	Tyr 485	Asp	Pro	Asn	Gly	Asn 490	Gln	Val	Asp	Tyr	Ser 495
Tyr	Thr	Ala	Tyr	Tyr 500	Gly	Phe	Glu	Lys	Val 505	Gly	Tyr	Tyr	Asn	Pro 510
Thr	Asp	Gly	Thr	Trp 515	Thr	Ile	Lys	Val	Val 520	Ser	Tyr	Ser	Gly	Ser 525
Ala	Asn	Tyr	Gln	Val 530	Asp	Val	Val	Ser	Asp 535	Gly	Ser	Leu	Ser	Gln 540
Pro	Gly	Ser	Ser	Pro 545	Ser	Pro	Gln	Pro	Glu 550	Pro	Thr	Val	Asp	Ala 555
Lys	Thr	Phe	Gln	Gly 560	Ser	Asp	His	Tyr	Tyr 565	Tyr	Asp	Arg	Ser	Asp 570
Thr	Phe	Thr	Met	Thr 575	Val	Asn	Ser	Gly	Ala 580	Thr	Lys	Ile	Thr	Gly 585
Asp	Leu	Val	Phe	Asp 590	Thr	Ser	Tyr	His	Asp 595	Leu	Asp	Leu	Tyr	Leu 600
Tyr	Asp	Pro	Asn	Gln 605	Lys	Leu	Val	Asp	Arg 610	Ser	Glu	Ser	Pro	Asn 615
Ser	Tyr	Glu	His	Val 620	Glu	Tyr	Leu	Thr	Pro 625	Ala	Pro	Gly	Thr	Trp 630
Tyr	Phe	Leu	Val	Tyr 635	Ala	Tyr	Tyr	Thr	Tyr 640	Gly	Trp	Ala	Tyr	Tyr 645
Glu	Leu	Thr	Ala	Lys 650	Val	Tyr	Tyr	Gly						

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TCTGAATTCG TTCTTTTCTG TATGG

25

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TGTACTGCTG GATCCGGCAG

20

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 80 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGATCCATCA GATTTTTGAG TGTAGATCAA CCAGTATGCT GCATTGTAA TTGTGAGATA 60

ATATCTCCCG CGGGTAAGGT 80

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

AGAGGCATGC GTATCCATCA GATTTTTGAG

30

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid

EST4453-04240

See  
Pg